

Group Theory, Linear Transformations, and Flows: (Some) Dynamical Systems on Manifolds

Moody T. Chu

North Carolina State University

presented at

Workshop on Lie Group Methods and Control Theory
June 30, 2004

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 03 JAN 2005		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Group Theory, Linear Transformations, and Flows: (Some) Dynamical Systems on Manifolds				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) North Carolina State University				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM001749, Lie Group Methods And Control Theory Workshop Held on 28 June 2004 - 1 July 2004., The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 47	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Outline

- Motivation
 - ◊ Realization Process
 - ◊ A Case Study
- Basic Form
 - ◊ Similarity Property
 - ◊ Decomposition Property
 - ◊ Reversal Property
- Matrix Groups and Group Actions
- Tangent Space and Projection
- Canonical Forms
- Objective Functions and Dynamical Systems
 - ◊ Examples
 - ◊ Least Squares
- New Thoughts
- Conclusion

Motivation

*What is the **simplest form** to which a family of matrices depending smoothly on the parameters can be reduced by a **change of coordinates** depending smoothly on the parameters?*

– V. I. Arnold

Geometric Methods in the Theory of Ordinary Differential Equations, 1988

- What is the simplest form referred to here?
- What kind of continuous change can be employed?

Realization Process

- Realization process, in a sense, means any deducible procedure that we use to rationalize and solve problems.
 - ◊ The simplest form refers to the agility to think and draw conclusions.
- In mathematics, a realization process often appears in the form of an iterative procedure or a differential equation.
 - ◊ The steps taken for the realization, i.e., the changes, could be discrete or continuous.

Continuous Realization

- Two abstract problems:
 - ◊ One is a make-up and is easy.
 - ◊ The other is the real problem and is difficult.
- A bridge:
 - ◊ A continuous path connecting the two problems.
 - ◊ A path that is easy to follow.
- A numerical method:
 - ◊ A method for moving along the bridge.
 - ◊ A method that is readily available.

Build the Bridge

- Specified guidance is available.
 - ◊ The bridge is constructed by monitoring the values of certain specified functions.
 - ◊ The path is guaranteed to work.
 - ◊ Such as the projected gradient method.
- Only some general guidance is available.
 - ◊ A bridge is built in a straightforward way.
 - ◊ No guarantee the path will be complete.
 - ◊ Such as the homotopy method.
- No guidance at all.
 - ◊ A bridge is built seemingly by accident.
 - ◊ Usually deeper mathematical theory is involved.
 - ◊ Such as the isospectral flows.

Characteristics of a Bridge

- A bridge, if it exists, usually is characterized by an ordinary differential equation.
- The discretization of a bridge, or a numerical method in travelling along a bridge, usually produces an iterative scheme.

Two Examples

-
-
- Eigenvalue Computation
 - Constrained Least Squares Approximation

The Eigenvalue Problem

- The mathematical problem:
 - ◊ A symmetric matrix A_0 is given.
 - ◊ Solve the equation

$$A_0 x = \lambda x$$

for a nonzero vector x and a scalar λ .

- An iterative method :
 - ◊ The QR decomposition:

$$A = QR$$

where Q is orthogonal and R is upper triangular.

- ◊ The QR algorithm (Francis'61):

$$\begin{aligned} A_k &= Q_k R_k \\ A_{k+1} &= R_k Q_k. \end{aligned}$$

- ◊ The sequence $\{A_k\}$ converges to a diagonal matrix.
- ◊ Every matrix A_k has the same eigenvalues of A_0 , i.e., $(A_{k+1} = Q_k^T A_k Q_k)$.

- A continuous method:

- ◊ Lie algebra decomposition:

$$X = X^o + X^+ + X^-$$

where X^o is the diagonal, X^+ the strictly upper triangular, and X^- the strictly lower triangular part of X .

- ◊ Define $\Pi_0(X) := X^- - X^{-\top}$.
- ◊ The Toda lattice (Symes'82, Deift et al'83):

$$\begin{aligned} \frac{dX}{dt} &= [X, \Pi_0(X)] \\ X(0) &= X_0. \end{aligned}$$

- ◊ Sampled at integer times, $\{X(k)\}$ gives the same sequence as does the QR algorithm applied to the matrix $A_0 = \exp(X_0)$.
- Evolution starts from X_0 and converges to the limit point of Toda flow, which is a diagonal matrix, maintains the spectrum.
 - ◊ The construction of the Toda lattice is based on the physics.
 - ▷ This is a Hamiltonian system.
 - ▷ A certain physical quantities are kept at constant, i.e., this is a *completely integrable* system.
 - ◊ The convergence is guaranteed by “nature”?

Least Squares Matrix Approximation

- The mathematical problem:
 - ◊ A symmetric matrix N and a set of real values $\{\lambda_1, \dots, \lambda_n\}$ are given.
 - ◊ Find a least squares approximation of N that has the prescribed eigenvalues.
- A standard formulation:

$$\begin{aligned} \text{Minimize } F(Q) &:= \frac{1}{2} \|Q^T \Lambda Q - N\|^2 \\ \text{Subject to } Q^T Q &= I. \end{aligned}$$

- ◊ Equality Constrained Optimization:
 - ▷ Augmented Lagrangian methods.
 - ▷ Sequential quadratic programming methods.
- ◊ None of these techniques is easy.
 - ▷ The constraint carries lots of redundancies.

- A continuous approach:
 - ◊ The projection of the gradient of F can easily be calculated.
 - ◊ Projected gradient flow (Brockiet'88, Chu&Driessel'90):

$$\begin{aligned} \frac{dX}{dt} &= [X, [X, N]] \\ X(0) &= \Lambda. \end{aligned}$$
- ▷ $X := Q^T \Lambda Q$.
 - ▷ Flow $X(t)$ moves in a descent direction to reduce $\|X - N\|^2$.
 - ◊ The optimal solution X can be fully characterized by the spectral decomposition of N and is unique.
- Evolution starts from an initial value and converges to the limit point, which solves the least squares problem.
 - ◊ The flow is built on the basis of systematically reducing the difference between the current position and the target position.
 - ◊ This is a descent flow.

Equivalence

- (Bloch'90) Suppose X is tridiagonal. Take

$$N = \text{diag}\{n, \dots, 2, 1\},$$

then

$$[X, N] = \Pi_0(X).$$

- A gradient flow hence becomes a Hamiltonian flow.

Basic Form

- Lax dynamics:

$$\begin{aligned}\frac{dX(t)}{dt} &:= [X(t), k_1(X(t))] \\ X(0) &:= X_0.\end{aligned}$$

- Parameter dynamics:

$$\begin{aligned}\frac{dg_1(t)}{dt} &:= g_1(t)k_1(X(t)) \\ g_1(0) &:= I.\end{aligned}$$

and

$$\begin{aligned}\frac{dg_2(t)}{dt} &:= k_2(X(t))g_2(t) \\ g_2(0) &:= I.\end{aligned}$$

$$\diamond k_1(X) + k_2(X) = X.$$

Similarity Property

- Define $Z(t) = g_1(t)X(t)g_1(t)^{-1}$.
- Check

$$X(t) = g_1(t)^{-1}X_0g_1(t) = g_2(t)X_0g_2(t)^{-1}.$$

$$\begin{aligned}\frac{dZ}{dt} &= \frac{dg_1}{dt}Xg_1^{-1} + g_1\frac{dX}{dt}g_1^{-1} + g_1X\frac{dg_1^{-1}}{dt} \\ &= (g_1k_1(X))Xg_1^{-1} \\ &\quad + g_1(Xk_1(X) - k_1(X)X)g_1^{-1} \\ &\quad + g_1X(-k_1(X)g_1^{-1}) \\ &= 0.\end{aligned}$$

- Thus $Z(t) = Z(0) = X(0) = X_0$.

Decomposition Property

- Trivially $\exp(X_0 t)$ satisfies the IVP

$$\exp(tX_0) = g_1(t)g_2(t).$$

$$\frac{dY}{dt} = X_0 Y, Y(0) = I.$$

- Define $Z(t) = g_1(t)g_2(t)$.
- Then $Z(0) = I$ and

$$\begin{aligned} \frac{dZ}{dt} &= \frac{dg_1}{dt}g_2 + g_1\frac{dg_2}{dt} \\ &= (g_1k_1(X))g_2 + g_1(k_2(X)g_2) \\ &= g_1Xg_2 \\ &= X_0Z \quad (\text{by Similarity Property}). \end{aligned}$$

- By the uniqueness theorem in the theory of ordinary differential equations, $Z(t) = \exp(X_0 t)$.

Reversal Property

$$\textcolor{red}{exp(tX(t))} = g_2(t)g_1(t).$$

- By Decomposition Property,

$$\begin{aligned} g_2(t)g_1(t) &= g_1(t)^{-1}exp(X_0t)g_1(t) \\ &= exp(g_1(t)^{-1}X_0g_1(t)t) \\ &= exp(X(t)t). \end{aligned}$$

Abstraction

- *QR*-type Decomposition:

- ◊ Lie algebra decomposition of $gl(n) \iff$ Lie group decomposition of $Gl(n)$ in the neighborhood of I .
- ◊ Arbitrary subspace decomposition $gl(n) \iff$ Factorization of a *one-parameter semigroup* in the neighborhood of I as the product of two nonsingular matrices , i.e.,

$$\textcolor{red}{exp(X_0 t) = g_1(t)g_2(t)}.$$

- ◊ The product $g_1(t)g_2(t)$ will be called the *abstract g_1g_2 decomposition* of $exp(X_0 t)$.

- *QR*-type Algorithm:

- ◊ By setting $t = 1$, we have

$$\begin{aligned} \textcolor{red}{exp(X(0))} &= g_1(1)g_2(1) \\ \textcolor{red}{exp(X(1))} &= g_2(1)g_1(1). \end{aligned}$$

- ◊ The dynamical system for $X(t)$ is autonomous \implies The above phenomenon will occur at every feasible integer time.
- ◊ Corresponding to the abstract g_1g_2 decomposition, the above iterative process for all feasible integers will be called the *abstract g_1g_2 algorithm*.

Matrix Groups

- A subset of nonsingular matrices (over any field) which are closed under matrix multiplication and inversion is called a *matrix group*.
 - ◊ Matrix groups are central in many parts of mathematics and applications.
- A smooth manifold which is also a group where the multiplication and the inversion are smooth maps is called a *Lie group*.
 - ◊ The most remarkable feature of a Lie group is that the structure is the same in the neighborhood of each of its elements.
- (Howe'83) Every (non-discrete) matrix group is in fact a Lie group.
 - ◊ Algebra and geometry are intertwined in the study of matrix groups.
- Lots of realization processes used in numerical linear algebra are the results of group actions.

Group	Subgroup	Notation	Characteristics
General linear		$Gl(n)$	$\{A \in \mathbb{R}^{n \times n} \mid \det(A) \neq 0\}$
	Special linear	$Sl(n)$	$\{A \in Gl(n) \mid \det(A) = 1\}$
Upper triangular		$\mathcal{U}(n)$	$\{A \in Gl(n) \mid A \text{ is upper triangular}\}$
	Unipotent	$\mathcal{Unip}(n)$	$\{A \in \mathcal{U}(n) \mid a_{ii} = 1 \text{ for all } i\}$
Orthogonal		$\mathcal{O}(n)$	$\{Q \in Gl(n) \mid Q^\top Q = I\}$
Generalized orthogonal		$\mathcal{O}_S(n)$	$\{Q \in Gl(n) \mid Q^\top S Q = S\}; \quad S \text{ is a fixed matrix}$
	Symplectic	$Sp(2n)$	$\mathcal{O}_J(2n); \quad J := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$
	Lorentz	$\mathcal{Lor}(n, k)$	$\mathcal{O}_L(n+k); \quad L := \text{diag}\{\underbrace{1, \dots, 1}_n, \underbrace{-1, \dots, -1}_k\}$
Affine		$\mathcal{A}ff(n)$	$\left\{ \left[\begin{array}{c c} A & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array} \right] \mid A \in Gl(n), \mathbf{t} \in \mathbb{R}^n \right\}$
	Translation	$\mathcal{T}rans(n)$	$\left\{ \left[\begin{array}{c c} I & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array} \right] \mid \mathbf{t} \in \mathbb{R}^n \right\}$
	Isometry	$\mathcal{I}som(n)$	$\left\{ \left[\begin{array}{c c} Q & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array} \right] \mid Q \in \mathcal{O}(n), \mathbf{t} \in \mathbb{R}^n \right\}$
Center of G		$Z(G)$	$\{z \in G \mid zg = gz, \text{ for every } g \in G\}, \quad G \text{ is a given group}$
Product of G_1 and G_2		$G_1 \times G_2$	$\{(g_1, g_2) \mid g_1 \in G_1, g_2 \in G_2\}; \quad (g_1, g_2) * (h_1, h_2) := (g_1 h_1, g_2 h_2); \quad G_1 \text{ and } G_2 \text{ are given groups}$
Quotient		G/N	$\{Ng \mid g \in G\}; \quad N \text{ is a fixed normal subgroup of } G$
	Hessenberg	$\mathcal{H}ess(n)$	$\mathcal{Unip}(n)/\mathcal{Z}_n$

Group Actions

- A function $\mu : G \times \mathbb{V} \longrightarrow \mathbb{V}$ is said to be a *group action* of G on a set \mathbb{V} if and only if
 - ◊ $\mu(gh, \mathbf{x}) = \mu(g, \mu(h, \mathbf{x}))$ for all $g, h \in G$ and $\mathbf{x} \in \mathbb{V}$.
 - ◊ $\mu(e, \mathbf{x}) = \mathbf{x}$, if e is the identity element in G .
- Given $\mathbf{x} \in \mathbb{V}$, two important notions associated with a group action μ :
 - ◊ The *stabilizer* of \mathbf{x} is

$$Stab_G(\mathbf{x}) := \{g \in G \mid \mu(g, \mathbf{x}) = \mathbf{x}\}.$$
 - ◊ The *orbit* of \mathbf{x} is

$$Orb_G(\mathbf{x}) := \{\mu(g, \mathbf{x}) \mid g \in G\}.$$

Set V	Group G	Action $\mu(g, A)$	Application
$\mathbb{R}^{n \times n}$	Any subgroup	$g^{-1}Ag$	conjugation
$\mathbb{R}^{n \times n}$	$\mathcal{O}(n)$	$g^\top Ag$	orthogonal similarity
$\underbrace{\mathbb{R}^{n \times n} \times \dots \times \mathbb{R}^{n \times n}}_k$	Any subgroup	$(g^{-1}A_1g, \dots, g^{-1}A_kg)$	simultaneous reduction
$S(n) \times S_{PD}(n)$	Any subgroup	$(g^\top Ag, g^\top Bg)$	symm. positive definite pencil reduction
$\mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n}$	$\mathcal{O}(n) \times \mathcal{O}(n)$	$(g_1^\top Ag_2, g_1^\top Bg_2)$	QZ decomposition
$\mathbb{R}^{m \times n}$	$\mathcal{O}(m) \times \mathcal{O}(n)$	$g_1^\top Ag_2$	singular value decomp.
$\mathbb{R}^{m \times n} \times \mathbb{R}^{p \times n}$	$\mathcal{O}(m) \times \mathcal{O}(p) \times \mathcal{G}l(n)$	$(g_1^\top Ag_3, g_2^\top Bg_3)$	generalized singular value decomp.

Some Exotic Group Actions (yet to be studied!)

- In numerical analysis, it is customary to use actions of the orthogonal group to perform the change of coordinates for the sake of cost efficiency and numerical stability.
 - ◊ What could be said if actions of the isometry group are used?
 - ▷ Being isometric, stability is guaranteed.
 - ▷ The inverse of an isometry matrix is easy.

$$\begin{bmatrix} Q & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} Q^\top & -Q^\top \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}.$$
 - ▷ The isometry group is larger than the orthogonal group.

- What could be said if actions of the orthogonal group plus shift are used?

$$\mu((Q, s), A) := Q^\top A Q + sI, \quad Q \in \mathcal{O}(n), s \in \mathbb{R}_+.$$

- What could be said if action of the orthogonal group with scaling are used?

$$\mu((Q, s), A) := sQ^\top A Q, \quad Q \in \mathcal{O}(n), s \in \mathbb{R}_+,$$

or

$$\mu((Q, \mathbf{s}, \mathbf{t}), A) := \text{diag}\{\mathbf{s}\} Q^\top A Q \text{diag}\{\mathbf{t}\}, \quad Q \in \mathcal{O}, \mathbf{s}, \mathbf{t} \in \mathbb{R}_+^n.$$

Tangent Space and Project Gradient

- Given a group G and its action μ on a set \mathbb{V} , the associated orbit $Orb_G(\mathbf{x})$ characterizes the rule by which \mathbf{x} is to be changed in \mathbb{V} .
 - ◊ Depending on the group G , an orbit is often too “wild” to be readily traced for finding the “simplest form” of \mathbf{x} .
 - ◊ Depending on the applications, a path/bridge/highway/differential equation needs to be built on the orbit to connect \mathbf{x} to its simplest form.
- A differential equation on the orbit $Orb_G(\mathbf{x})$ is equivalent to a differential equation on the group G .
 - ◊ Lax dynamics on $X(t)$.
 - ◊ Parameter dynamics on $g_1(t)$ or $g_2(t)$.
- To stay in either the orbit or the group, the vector field of the dynamical system must be distributed in the tangent space of the corresponding manifold.
- Most of the tangent spaces for the matrix groups can be calculated explicitly.
- If some kind of objective function has been used to control the connecting bridge, its gradient should be projected to the tangent space.

Tangent Space in General

- Given a matrix group $G \leq \mathcal{GL}(n)$, the *tangent space* to G at $A \in G$ can be defined as

$$\mathcal{T}_A G := \{\gamma'(0) \mid \gamma \text{ is a differentiable curve in } G \text{ with } \gamma(0) = A\}.$$

- The tangent space $\mathfrak{g} = \mathcal{T}_I G$ at the identity I is critical.

◊ \mathfrak{g} is a Lie subalgebra in $\mathbb{R}^{n \times n}$, i.e.,

If $\alpha'(0), \beta'(0) \in \mathfrak{g}$, then $[\alpha'(0), \beta'(0)] \in \mathfrak{g}$

- ◊ The tangent space of a matrix group has the same structure everywhere, i.e.,

$$\mathcal{T}_A G = A\mathfrak{g}.$$

- ◊ $\mathcal{T}_I G$ can be characterized as the *logarithm* of G , i.e.,

$$\mathfrak{g} = \{M \in \mathbb{R}^{n \times n} \mid \exp(tM) \in G, \text{ for all } t \in \mathbb{R}\}.$$

Group G	Algebra \mathfrak{g}	Characteristics
$Gl(n)$	$gl(n)$	$\mathbb{R}^{n \times n}$
$Sl(n)$	$sl(n)$	$\{M \in gl(n) \text{trace}(M) = 0\}$
$Aff(n)$	$aff(n)$	$\left\{ \begin{bmatrix} M & \mathbf{t} \\ \mathbf{0} & 0 \end{bmatrix} \mid M \in gl(n), \mathbf{t} \in \mathbb{R}^n \right\}$
$O(n)$	$o(n)$	$\{K \in gl(n) K \text{ is skew-symmetric}\}$
$Isom(n)$	$isom(n)$	$\left\{ \begin{bmatrix} K & \mathbf{t} \\ \mathbf{0} & 0 \end{bmatrix} \mid K \in o(n), \mathbf{t} \in \mathbb{R}^n \right\}$
$G_1 \times G_2$	$\mathcal{T}_{(e_1, e_2)} G_1 \times G_2$	$\mathfrak{g}_1 \times \mathfrak{g}_2$

An Illustration of Projection

- The tangent space of $\mathcal{O}(n)$ at any orthogonal matrix Q is

$$\mathcal{T}_Q \mathcal{O}(n) = Q\mathbb{K}(n)$$

where

$$\mathbb{K}(n) = \{\text{All skew-symmetric matrices}\}.$$

- The normal space of $\mathcal{O}(n)$ at any orthogonal matrix Q is

$$\mathcal{N}_Q \mathcal{O}(n) = Q\mathbb{S}(n).$$

- The space $\mathbb{R}^{n \times n}$ is split as

$$\mathbb{R}^{n \times n} = Q\mathbb{S}(n) \oplus Q\mathbb{K}(n).$$

- A unique orthogonal splitting of $X \in \mathbb{R}^{n \times n}$:

$$X = Q(Q^T X) = Q \left\{ \frac{1}{2}(Q^T X - X^T Q) \right\} + Q \left\{ \frac{1}{2}(Q^T X + X^T Q) \right\}.$$

- The projection of X onto the tangent space $\mathcal{T}_Q \mathcal{O}(n)$ is given by

$$\text{Proj}_{\mathcal{T}_Q \mathcal{O}(n)} X = Q \left\{ \frac{1}{2}(Q^T X - X^T Q) \right\}.$$

Canonical Forms

- A canonical form refers to a “specific structure” by which a certain conclusion can be drawn or a certain goal can be achieved.
- The superlative adjective “simplest” is a relative term which should be interpreted broadly.
 - ◊ A matrix with a specified pattern of zeros, such as a diagonal, tridiagonal, or triangular matrix.
 - ◊ A matrix with a specified construct, such as Toeplitz, Hamiltonian, stochastic, or other linear varieties.
 - ◊ A matrix with a specified algebraic constraint, such as low rank or nonnegativity.

Canonical form	Also know as	Action
Bidiagonal J	Quasi-Jordan Decomp., $A \in \mathbb{R}^{n \times n}$	$P^{-1}AP = J,$ $P \in \mathcal{G}l(n)$
Diagonal Σ	Sing. Value Decomp., $A \in \mathbb{R}^{m \times n}$	$U^T AV = \Sigma,$ $(U, V) \in \mathcal{O}(m) \times \mathcal{O}(n)$
Diagonal pair (Σ_1, Σ_2)	Gen. Sing. Value Decomp., $(A, B) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{p \times n}$	$(U^T AX, V^T BX) = (\Sigma_1, \Sigma_2),$ $(U, V, X) \in \mathcal{O}(m) \times \mathcal{O}(p) \times \mathcal{G}l(n)$
Upper quasi-triangular H	Real Schur Decomp., $A \in \mathbb{R}^{n \times n}$	$Q^T AQ = H,$ $Q \in \mathcal{O}(n)$
Upper quasi-triangular H Upper triangular U	Gen. Real Schur Decomp., $A, B \in \mathbf{R}^{n \times n}$	$(Q^T AZ, Q^T BZ) = (H, U),$ $Q, Z \in \mathcal{O}(n)$
Symmetric Toeplitz T	Toeplitz Inv. Eigenv. Prob., $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{R}$ is given	$Q^T \text{diag}\{\lambda_1, \dots, \lambda_n\}Q = T,$ $Q \in \mathcal{O}(n)$
Nonnegative $N \geq 0$	Nonneg. inv. Eigenv. Prob., $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{C}$ is given	$P^{-1} \text{diag}\{\lambda_1, \dots, \lambda_n\}P = N,$ $P \in \mathcal{G}l(n)$
Linear variety X with fixed entries at fixed locations	Matrix Completion Prob., $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{C}$ is given $X_{i_\nu, j_\nu} = a_\nu, \nu = 1, \dots, \ell$	$P^{-1}\{\lambda_1, \dots, \lambda_n\}P = X,$ $P \in \mathcal{G}l(n)$
Nonlinear variety with fixed singular values and eigenvalues	Test Matrix Construction, $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ and $\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_n\}$ are given	$P^{-1}\Lambda P = U^T \Sigma V$ $P \in \mathcal{G}l(n), \quad U, V \in \mathcal{O}(n)$
Maximal fidelity	Structured Low Rank Approx. $A \in \mathbb{R}^{m \times n}$	$(\text{diag}(USS^T U^T))^{-1/2} USV^T,$ $(U, S, V) \in \mathcal{O}(m) \times \mathbb{R}_x^k \times \mathcal{O}(n)$

Objective Functions

- The orbit of a selected group action only defines the rule by which a transformation is to take place.
- Properly formulated objective functions helps to control the construction of a bridge between the current point and the desired canonical form on a given orbit.
 - ◊ The bridge often assumes the form of a differential equation on the manifold.
 - ◊ The vector field of the differential equation must distributed over the tangent space of the manifold.
 - ◊ Corresponding to each differential equation on the orbit of a group action is a differential equation on the group, and vice versa.
- **How to choose appropriate objective functions?**

Some Flows on $Orb\mathcal{O}_{(n)}(X)$ under Conjugation

- Toda lattice arises from a special mass-spring system (Symes'82, Deift et al'83),

$$\begin{aligned} \frac{dX}{dt} &= [X, \Pi_0(X)], & \Pi_0(X) &= X^{-\top} - X^{-\top}, \\ X(0) &= \text{tridiagonal and symmetric.} \end{aligned}$$

- ◊ No specific objective function is used.
 - ▷ Physics law governs the definition of the vector field.
- ◊ Generalization to general matrices is totally by brutal force and blindness (*and by the then young and desperate researchers*) (Chu'84, Watkins'84).

$$\frac{dX}{dt} = [X, \Pi_0(G(X))], \quad G(z) \text{ is analytic over spectrum of } X(0).$$
 - ▷ But nicely explains the pseudo-convergence and convergence behavior of the classical QR algorithm for general and normal matrices, respectively.
 - ▷ Sorting of eigenvalues at the limit point is observed, but not quite clearly understood.

- Double bracket flow (Brockett'88),

$$\frac{dX}{dt} = [X, [X, N]], \quad N = \text{fixed and symmetric.}$$
 - ◊ This is the projected gradient flow of the objective function

$$\begin{aligned} \text{Minimize } F(Q) &:= \frac{1}{2} \|Q^T \Lambda Q - N\|^2, \\ \text{Subject to } Q^T Q &= I. \end{aligned}$$
 - ▷ Sorting is necessary in the first order optimality condition (Wielandt&Hoffman'53).
- Take a special $N = \text{diag}\{n, n-1, \dots, 2, 1\}$,
 - ◊ X is tridiagonal and symmetric \implies Double bracket flow \equiv Toda lattice (Bloch'90).
 - ▷ **Bingo!** The classical Toda lattice does have an objective function in mind.
 - ◊ X is a general symmetric matrix \implies Double bracket = A specially scaled Toda lattice.
- Scaled Toda lattice (Chu'95),

$$\frac{dX}{dt} = [X, K \circ X], \quad K = \text{fixed and skew-symmetric.}$$
 - ◊ Flexible in componentwise scaling.
 - ◊ Enjoy very general convergence behavior.
 - ◊ But still no explicit objective function in sight.

Some Flows on $Orb_{\mathcal{O}(m) \times \mathcal{O}(n)}(X)$ under Equivalence

- Any flow on the orbit $Orb_{\mathcal{O}(m) \times \mathcal{O}(n)}(X)$ under equivalence must be of the form

$$\frac{dX}{dt} = X(t)h(t) - k(t)X(t), \quad h(t) \in \mathbb{K}(n), \quad k(t) \in \mathbb{K}(m).$$

- QZ flow (Chu'86),

$$\begin{aligned} \frac{dX_1}{dt} &= X_1 \Pi_0(X_2^{-1} X_1) - \Pi_0(X_1 X_2^{-1}) X_1, \\ \frac{dX_2}{dt} &= X_2 \Pi_0(X_2^{-1} X_1) - \Pi_0(X_1 X_2^{-1}) X_2, \end{aligned}$$

- SVD flow (Chu'86),

$$\begin{aligned} \frac{dY}{dt} &= Y \Pi_0(Y(t)^\top Y(t)) - \Pi_0(Y(t)Y(t)^\top) Y, \\ Y(0) &= \text{bidagonal}. \end{aligned}$$

- ◊ The "objective" in the design of this flow was to maintain the bidiagonal structure of $Y(t)$ for all t .
- ◊ The flow gives rise to the Toda flows for $Y^\top Y$ and $Y Y^\top$.

Projected Gradient Flows

- Given
 - ◊ A continuous matrix group $G \subset \mathcal{GL}(n)$.
 - ◊ A fixed $X \in \mathbb{V}$ where $\mathbb{V} \subset \mathbb{R}^{n \times n}$ be a subset of matrices.
 - ◊ A differentiable map $f : \mathbb{V} \rightarrow \mathbb{R}^{n \times n}$ with a certain “inherent” properties, e.g., symmetry, isospectrum, low rank, or other algebraic constraints.
 - ◊ A group action $\mu : G \times \mathbb{V} \rightarrow \mathbb{V}$.
 - ◊ A projection map P from $\mathbb{R}^{n \times n}$ onto a singleton, a linear subspace, or an affine subspace $\mathbb{P} \subset \mathbb{R}^{n \times n}$ where matrices in \mathbb{P} carry a certain desired structure, e.g., the canonical form.
- Consider the functional $F : G \rightarrow \mathbb{R}$

$$F(g) := \frac{1}{2} \|f(\mu(g, X)) - P(\mu(g, X))\|_F^2.$$
- Want to minimize F over G .
- Flow approach:
 - ◊ Compute $\nabla F(g)$.
 - ◊ Project $\nabla F(g)$ onto $\mathcal{T}_g G$.
 - ◊ Follow the projected gradient until convergence.

Some Old Examples

- Brockett's double bracket flow (Brockett'88).
- Least squares approximation with spectral constraints (Chu&Driessel'90).

$$\frac{dX}{dt} = [X, [X, P(X)]].$$

- Simultaneous reduction problem (Chu'91),

$$\begin{aligned} \frac{dX_i}{dt} &= \left[X_i, \sum_{j=1}^p \frac{[X_j, P_j^T(X_j)] - [X_j, P_j^T(X_j)]^T}{2} \right] \\ X_i(0) &= A_i \end{aligned}$$

- Nearest normal matrix problem (Chu'91),

$$\begin{aligned} \frac{dW}{dt} &= \left[W, \frac{1}{2} \{ [W, \text{diag}(W^*)] - [W, \text{diag}(W^*)]^* \} \right] \\ W(0) &= A. \end{aligned}$$

- Matrix with prescribed diagonal entries and spectrum (Schur-Horn Theorem) (Chu'95),

$$\dot{X} = [X, [\text{diag}(X) - \text{diag}(a), X]]$$

- Inverse generalized eigenvalue problem for symmetric-definite pencil (Chu&Guo'98).

$$\dot{X} = -((XW)^T + XW),$$

$$\dot{Y} = -((YW)^T + YW),$$

$$W := X(X - P_1(X)) + Y(Y - P_2(Y)).$$

- Various structured inverse eigenvalue problems (Chu&Golub'02).
- Remember the list of applications that Nicoletta gave on Monday!!!!

New Thoughts

- The idea of group actions, least squares, and the corresponding gradient flows can be generalized to other structures such as
 - ◊ Stiefel manifold $\mathcal{O}(p, q) := \{Q \in \mathbb{R}^{p \times q} \mid Q^T Q = I_q\}$.
 - ◊ The manifold of oblique matrices $\mathcal{OB}(n) := \{Q \in \mathbb{R}^{n \times n} \mid \text{diag}(Q^T Q) = I_n\}$.
 - ◊ Cone of nonnegative matrices.
 - ◊ Semigroups.
 - ◊ Low rank approximation.
- Using the product topology to describe separate groups and actions might broaden the applications.
- Any advantages of using the isometry group over the orthogonal group?

Stochastic Inverse Eigenvalue Problem

-
-
- Construct a stochastic matrix with prescribed spectrum
 - ◊ A hard problem (Karpelevic'51, Minc'88).

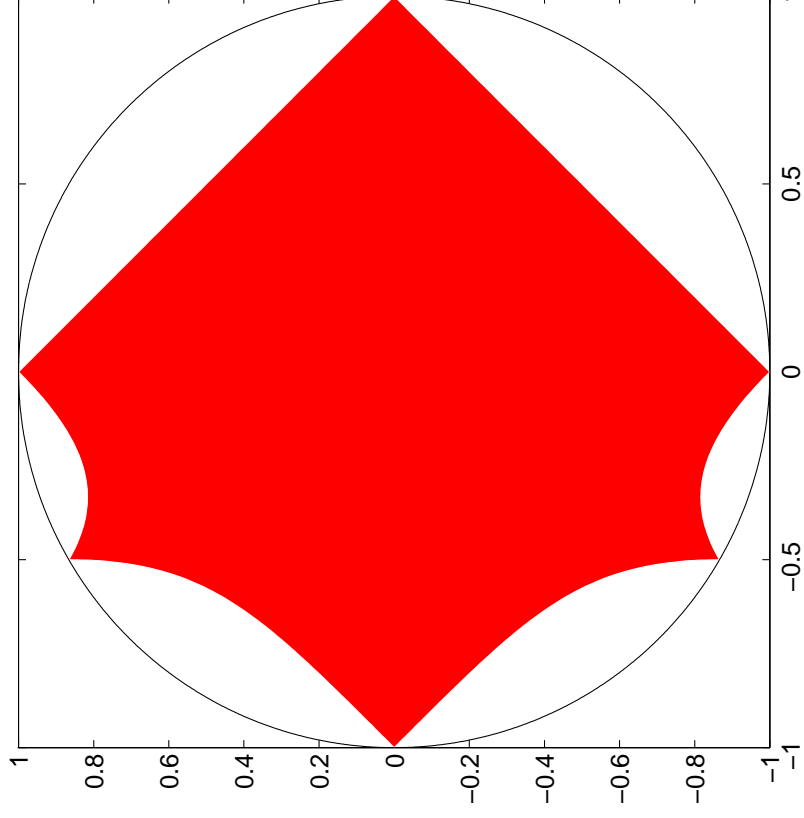


Figure 1: Θ_4 by the Karpelevič theorem.

- ◊ Would be done if the nonnegative inverse eigenvalue problem is solved – a long standing open question.

- Least squares formulation:

$$\begin{array}{ll} \text{Minimize} & F(g, R) := \frac{1}{2} \|gJg^{-1} - R \circ R\|^2 \\ \text{Subject to} & g \in Gl(n), \ R \in gl(n). \end{array}$$

- ◊ J = Real matrix carrying spectral information.
- ◊ \circ = Hadamard product.

- Steepest descent flow:

$$\begin{array}{ll} \frac{dg}{dt} & = \ [(gJg^{-1})^T, \alpha(g, R)] g^{-T} \\ \frac{dR}{dt} & = \ 2\alpha(g, R) \circ R. \end{array}$$

- ◊ $\alpha(g, R) := gJg^{-1} - R \circ R$.

- ASVD flow for g (Bunse-Gerstner et al'91, Wright'92):

$$\begin{aligned} g(t) &= X(t)S(t)Y(t)^T \\ \dot{g} &= \dot{X}SY^T + X\dot{S}Y^T + X\dot{S}\dot{Y}^T \\ X^T \dot{g} Y &= \underbrace{X^T \dot{X}}_Z S + \dot{S} + S \underbrace{\dot{Y}^T}_W Y \end{aligned}$$

Define $Q := X^T \dot{g} Y$. Then

$$\begin{aligned} \frac{dS}{dt} &= \text{diag}(Q), \\ \frac{dX}{dt} &= XZ, \\ \frac{dY}{dt} &= YW. \end{aligned}$$

◊ Z, W are skew-symmetric matrices obtainable from Q and S .

Nonnegative Matrix Factorization

- For various applications, given a nonnegative matrix $A \in \mathbb{R}^{m \times n}$, want to

$$\min_{0 \leq V \in \mathbb{R}^{m \times k}, 0 \leq H \in \mathbb{R}^{k \times n}} \frac{1}{2} \|A - VH\|_F^2.$$

- ◊ Relatively new techniques for dimension reduction applications.

- ▷ Image processing — no negative pixel values.

- ▷ Data mining — no negative frequencies.

- ◊ No firm theoretical foundation available yet (Tropp'03).

- Relatively easy by flow approach!

$$\min_{E \in \mathbb{R}^{m \times k}, F \in \mathbb{R}^{k \times n}} \frac{1}{2} \|A - (E \circ E)(F \circ F)\|_F^2.$$

- Gradient flow:

$$\begin{aligned} \frac{dV}{dt} &= V \circ (A - VH)H^\top, \\ \frac{dH}{dt} &= H \circ (V^\top(A - VH)). \end{aligned}$$

- ◊ Once any entry of either V or H hits 0, it stays zero. This is a natural barrier!
 - ◊ The first order optimality condition is clear.

Image Articulation Library

- Assume images are composite objects in many articulations and poses.
- Factorization would enable the identification and classification of intrinsic “parts” that make up the object being imaged by multiple observations.
- Each column \mathbf{a}_j of a nonnegative matrix A now represents m pixel values of one image.
- The columns \mathbf{v}_k of V are k basis elements in \mathbb{R}^m .
- The columns of H , belonging to \mathbb{R}^k , can be thought of as coefficient sequences representing the n images in the basis elements.

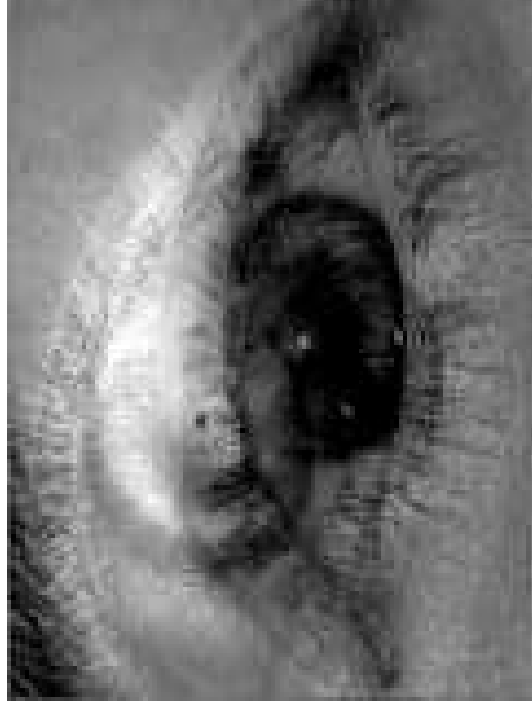
$A \in \mathbb{R}^{19200 \times 10}$ Representing 10 Gray-scale 120×160 Irises



Basis Irises with $k = 2$



(Wrong?) Basis Irises with $k = 4$



Conclusion

- Many operations used to transform matrices can be considered as matrix group actions.
- The view unifies different transformations under the same framework of tracing orbits associated with corresponding group actions.
 - ◊ More sophisticated actions can be composed that might offer the design of new numerical algorithms.
 - ◊ As a special case of Lie groups, (tangent space) structure of a matrix group is the same at every of its element. Computation is easy and cheap.
- It is yet to be determined how a dynamical system should be defined over a group so as to locate the simplest form.
 - ◊ The notion of “simplicity” varies according to the applications.
 - ◊ Various objective functions should be used to control the dynamical systems.
 - ◊ Usually offers a global method for solving the underlying problem.
- Continuous realization methods often enable to tackle existence problems that are seemingly impossible to be solved by conventional discrete methods.
- Group actions together with properly formulated objective functions can offer a channel to tackle various classical or new and challenging problems.

- Some basic ideas and examples have been outlined in this talk.
 - ◊ More sophisticated actions can be composed that might offer the design of new numerical algorithms.
 - ◊ The list of application continues to grow.
- New computational techniques for structured dynamical systems on matrix group will further extend and benefit the scope of this interesting topic.
 - ◊ Need ODE techniques specially tailored for gradient flows.
 - ◊ Need ODE techniques suitable for very large-scale dynamical systems.
 - ◊ **Help! Help! Help!**